

# Scene Identification Using Discriminative Pattern

Hillol Chakraborty  
111CS0202



Department of Computer Science  
National Institute of Technology, Rourkela  
May, 2015

# Scene Identification Using Discriminative Pattern

*Thesis submitted in partial fulfillment  
of the requirements for the degree of*

Bachelor of Technology  
*in*  
Computer Science and Engineering

*by*

Hillol Chakraborty  
111CS0202

*Under the supervision of*

Prof. Ramesh Kumar Mohapatra



Department of Computer Science  
National Institute of Technology, Rourkela  
May, 2015



Department of Computer Science and Engineering  
**National Institute of Technology, Rourkela**  
Rourkela-769008, Odisha, India

May, 2015

## Certificate

This is to certify that the work in the project entitled *Scene Identification using Discriminative Pattern* by *Hillol Chakraborty* is a record of his work carried out under my supervision and guidance in partial fulfillment of the requirements for the award of the degree of *Bachelor of Technology* in *Computer Science and Engineering*.

Prof. Ramesh Kumar Mohapatra,  
Assistant Professor

# Acknowledgement

I would like to express my special appreciation and thanks to my project supervisor Prof. Ramesh Kr Mohapatra, you have been an tremendous mentor for me. Your advice on both research and career has been priceless for me. I would like to thank out department's HOD, Prof S. K. Rath and all other faculty members of our department who have always been a great source of knowledge and inspiration for me.

I would also like to extend my thanks to my parents and family members for their great support without which I could have never done my project successfully. Last but not the least I thank my fellow friends and batch mates who have always been on my side during all the ups and downs of my work.

# Abstract

Scene Identification is the process of retrieving relevant information about an image. This can be used in modern life in various ways. The image clicked by a mobile camera can be input to a remote server which returns some relevant information about the image or what we call as scene. In this project we have worked on a method for scene identification using discriminative patterns.

We rank the image patches sampled from target images. For classification purposes, we have used Support Vector Classifiers. SVMs are trained using top discriminative patches of image classes. We vary number of top discriminative patches and finally get the number of discriminative patterns which yields the best result. For demonstration, we have used ZuBud(Zurich Building dataset) as well as NIT Rourkela Building dataset.

# Contents

<b>Acknowledgement</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Scene Identification . . . . .	1
1.2 Why discriminative measure? . . . . .	1
<b>2 Literature Review</b>	<b>3</b>
<b>3 Scene Identification</b>	<b>5</b>
<b>4 Proposed Method</b>	<b>6</b>
4.1 Dataset Used . . . . .	6
4.2 Block Diagram . . . . .	7
4.3 Preprocessing . . . . .	7
4.4 Feature Extraction . . . . .	8
4.4.1 Color Feature . . . . .	8
4.4.2 Texture Feature . . . . .	8
4.5 Ranking Image Patches . . . . .	9
4.6 Training SVM by top discriminative patterns . . . . .	10
<b>5 Results and Conclusions</b>	<b>12</b>
5.1 Result without k-means Clustering . . . . .	12

5.2 K-Means Clustering Result . . . . .	13
---	----

5.3 Conclusion . . . . .	14
--------------------------	----

<b>References</b>	<b>16</b>
-------------------	-----------

# Chapter 1

## Introduction

### 1.1 Scene Identification

Scene Identification is the process of retrieving relevant information about an image. Our visual system can gather incredible amount of information about an image at a glance. We have named different names of scenes such as “building x”, “beach”, “animal” etc. So, scene identification can be reduced to get the scene name out of an input image.

This scene identification is a very developing area of research today because its result can be used in various ways. It can be implemented in maps as well. The photo clicked by a mobile camera can be used as an input to a remote server which returns the name of the scene contained in the image. While contextual scene identification has already been worked on but unconstrained scene identification is still a challenge.

### 1.2 Why discriminative measure?

Our approach is rather intuitive. When we look at an image and recognize the scene on it, what we actually do is we look at some particular portions of the image that helps us distinguish it from other images. This what we call discriminative patterns. If certain pattern is repeatedly occurring in



a particular class of image then this is called the discriminative pattern of that image class. So instead of using the complete image for classification we can rather select top discriminative portions of the image and use it for classification. This helps us avoid occlusions like unwanted trees, ground or some persons standing in front of the scene. Invariant local descriptors of image patches that are extracted around interest points detected in an image for image matching is a very efficient approach. The reason is that it represents a visual entity as its parts and allows flexible modeling of the geometrical relation among the parts. Its specialty is to focus on that part of image which are particular to the class. It is pretty obvious that this part-based approach can handle scenes with occlusions.

The organization of the thesis is as follows: The first chapter is a brief introduction about the literature review. The second chapter is a brief introduction about Scene Identification, its various uses and implementation. The third chapter describes the method implemented in this project. It has description about the data set used, block diagram representation of the whole project etc. The fourth chapter shows different results that we got finally. It has also the comparison between the two methods that we implemented and shows which one is better.

# Chapter 2

## Literature Review

Eric Nowak et al.[1] have proposed a method for image sampling strategies for bag of features image classification. The basic idea is that image is considered as a loose collection of independent patches, sampling a representative set of patches from that image, evaluating a visual descriptor vector for each patch independently and using the resulting distribution of sampling descriptor space as a characterization of image.

Frederic Jurie et al.[2] have worked on creating efficient codebooks for visual recognition. Codebooks are generally created with the help of k-means clustering from a set of training images. In their paper, they have shown that in case of dense sampling, k-means over-adapts to this. clustering centers are mostly around densest regions.

Li Fei-Fei et al.[3] have prepared a Bayesian hierarchical model for learning natural scene catagories. They have worked by representing images of a scene as collection of local regions, denoted as codewords obtained by unsupervised learning.

Joo-Hwee Lim et al.[4] have published their work as scene identification using discriminative patterns. Here they have focussed on the fact that not the whole image, but some portions of it actually distinguish it from images of other classes. Supervised SVM has been used for classification in this case.

Thomas Deselaers et al.[5] have devised an algorithm on discriminative

training for object recognition using image patches. The approach applies discriminative training of log-linear models to image patch histograms.

Dong ping Tian et al.[6] have reviewed on image feature extraction and representation techniques. They focused on global and local feature of image feature extraction techniques.

# Chapter 3

## Scene Identification

Scene Identification is the process of retrieving relevant information about an image. We have named different names of scenes such as “building x”, “beach”, “animal” etc. This scene identification is a very developing area of research today because its result can be used in various ways. It can be implemented in maps as well. The photo clicked by a mobile camera can be used as an input to a remote server which returns the name of the scene contained in the image. While contextual scene identification has already been worked on but unconstrained scene identification is still a challenge.

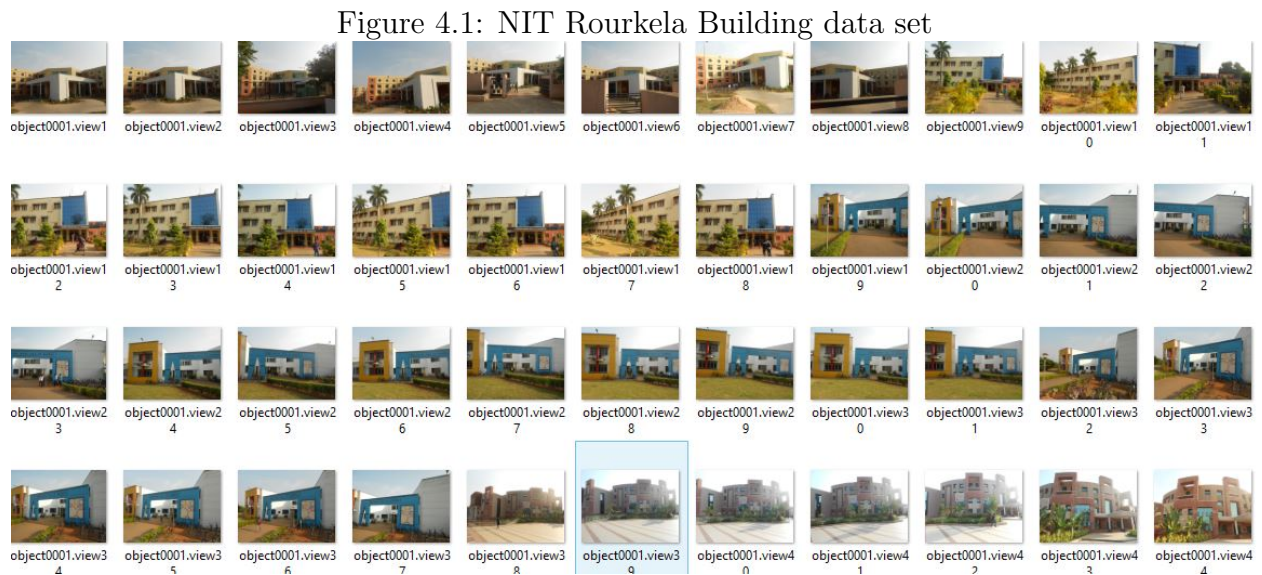
Jitendra Malik et al have worked on scene identification in the paper “Why scene identification is just Texture Analysis?” In the paper, it is proposed that a simple texture analysis can help in identifying the scene in an image. The method learns texture in different parts of the image and finally uses the knowledge to identify new scenes. The texture analysis leads to similar identifications and confusions as subjects with limited processing time.

# Chapter 4

## Proposed Method

### 4.1 Dataset Used

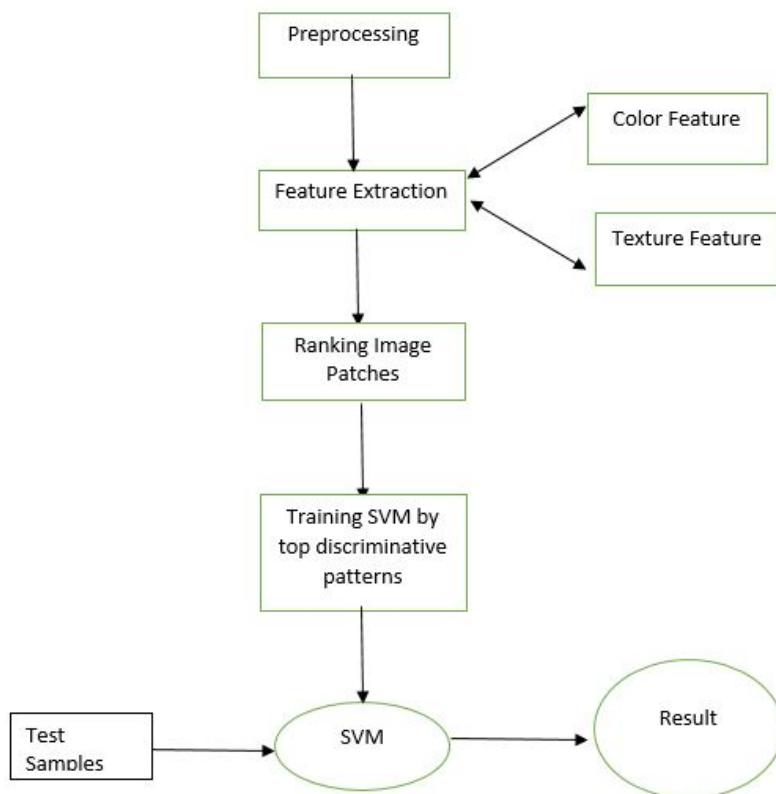
For the purpose of our demonstration, we have used NIT Rourkela building data set. We have selected 8-10 buildings in NIT Rourkela campus and have taken around ten photos of each building from different angles. Different angles have led to inclusion of sky, ground and even occlusions like people and trees etc. 80 percent of the images have been chosen for training the system and the rest 20 percent is left for testing purposes. A glimpse of the dataset is given in the following figure:



## 4.2 Block Diagram

Block diagram of the whole process is shown in Figure 4.2.

Figure 4.2: Block Diagram of the whole process



## 4.3 Preprocessing

All the images are resized to  $240 \times 320$  for standardization purpose. We take each image from the training set and divide them into overlapping patches of  $60 \times 60$  dimensions after every 20pixels in both horizontal and vertical directions. So, we get 140 patches out of a single  $240 \times 320$  image. The following figure shows that:

Figure 4.3: Image Sampling



## 4.4 Feature Extraction

We have employed two kinds of features here:

- Color Feature
- Texture Feature

### 4.4.1 Color Feature

Color feature can be very informative here because we are dealing with colored images here. For color feature extraction, we have used color histogram properties. We have taken mean and standard deviation of each R, G and B histograms leading to a feature vector of length 6.

### 4.4.2 Texture Feature

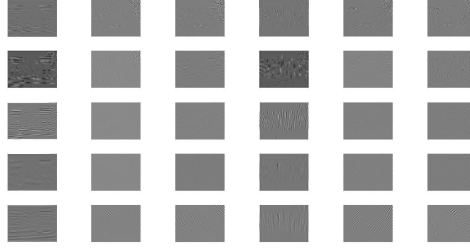
For texture feature extraction, we have used Gabor Transformation here. We have filtered each image patch with gabor filters of 5 phases and 6 orienta-

tions. After filtering, we take the mean and standard deviation from each filtered image thereby making a feature vector of length  $(5 * 6 * 2 = 60)$ . A figure showing an image filtered by gabor filters:

Figure 4.4: Gabor Filterd image magnitude



Figure 4.5: Gabor Filterd image real parts



## 4.5 Ranking Image Patches

The main point of this methodology is "Discriminative Patterns" which means that certain patterns which appear frequently in a particular class and doesn't occur frequently in other classes. These discriminative patterns define the classes because they are particular to their own class. Measuring discriminative classes is explained below.

We calculate probability of occurrence of a patch in its own class as  $P(z_c)$  and its probability of occurrence in union of all other classes is  $P(z_{c'})$ . We call a patch discriminative if its probability of occurrence in its own class is



more and in other classes is less. So, we calculate likelihood as the following:

$$L(z) = P(z_c)/P(z_{c'})$$

Now for calculating probability density function,  $P(z)$ , we have used non parametric probability density estimator Parzen Window with a Gaussian Kernal.

$$P(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{(h\sqrt{2\pi})^d} \exp\left(-\frac{1}{2}\left(\frac{x-x_i}{h}\right)^2\right),$$

where  $h$  is the standard deviation of the Gaussian PDF along each dimension.

It is very intuitive that matter of importance generally present at the center of the image. So we define a weight function that gives weight to the portions near the center.

$$W(z) = \frac{1}{2\pi} \exp^{-\frac{1}{2}[(x-x_c)^2+(y-y_c)^2]}$$

where  $x, y$  are the coordinates of the center of the patch and  $x_c$  and  $y_c$  are the center of the whole image. Then, we rank image patches  $z$  based on  $L(z) * W(z)$ . Next, we can take any number of top ranked patches for training purpose.

## 4.6 Training SVM by top discriminative patterns

SVM is a supervised classifier. We used multiSVM to classify all the sample points into classes. There are lots of kernel functions that can be used in SVM. For our purpose, we have employed RBF kernal function. RBF kernal for two vectors  $\mathbf{x}$  and  $\mathbf{x}'$  is represented as

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right)$$

Figure 4.6: Taking top 70 patches



where,  $\|\mathbf{x} - \mathbf{x}'\|^2$  is squared euclidean distance between two feature vectors. Since in Matlab, function for binary SVM is given, so we have written an algorithm to extend it to multiSVM. Th algorithm is as follows:

```

multiSVM(trainingSet,group){
u=unique(group);
numberClasses=length(u);
for(k=1:numberClasses){
g=(group==u(k))
svm(k) = svmtrain(trainingSet,g);
}}

```

# Chapter 5

## Results and Conclusions

### 5.1 Result without k-means Clustering

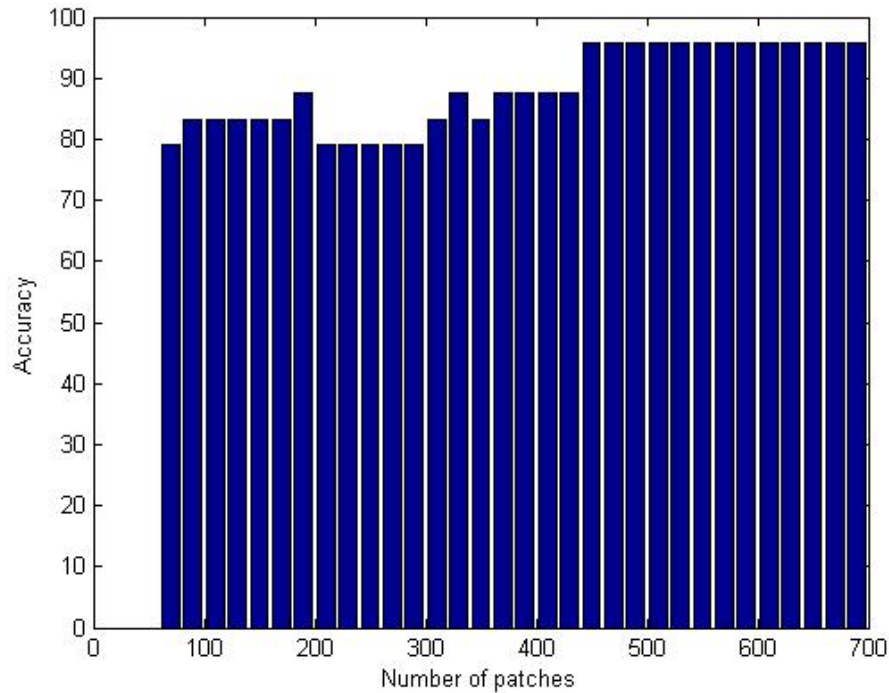
In this thesis, we have worked on a procedure for scene identification. Our approach is using Discriminative Patterns. For demonstration, we have used NIT Rourkela Building dataset. We have taken 80% of the photos for training and 20% are kept for testing. In discriminative measure, we have ranked image patches of each class and taken different number of patches different times to train our system and measure accuracy. Below is a table 5.1 given showing number of top discriminative patches versus accuracy:

So, it is clear that increasing number of patches for training increases the accuracy of scene identification. But, increasing number of patches also means increasing processing time. So, it is preferable to select an intermediate number of patches like 200 is good enough.

Table 5.1: Table for No. of patches vs accuracy

Number of Patches	Accuracy
70	79.1667%
90	83.3333%
110	83.3333%
130	83.3333%
150	83.3333%
170	87.5000%
190	87.5000%

Figure 5.1: Result on NIT dataset

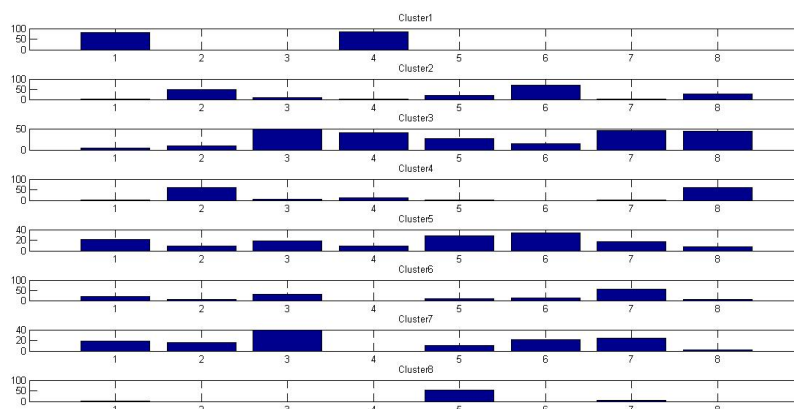


## 5.2 K-Means Clustering Result

We made eight clusters out of training set and formed multiSVMs for each cluster. The cluster's mean which is closest to the test data is considered suitable for that test data. Finally, the test data is tested with the SVM of that particular cluster to get the result. The cluster distribution is shown in

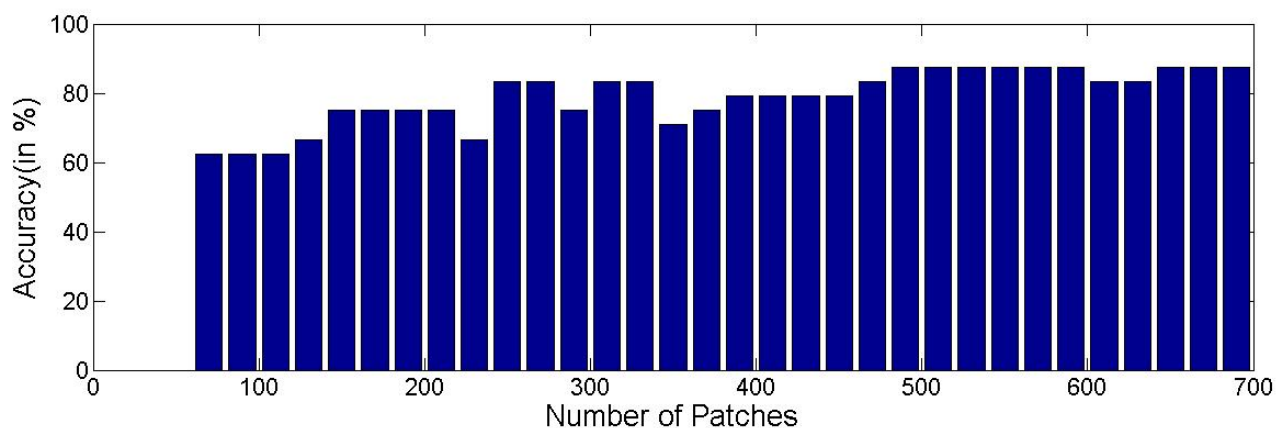
figure 5.2

Figure 5.2: Cluster Distribution



The efficiency bar graph is shown in figure 5.3

Figure 5.3: Result with K-means clustering



### 5.3 Conclusion

We have implemented scene identification using discriminative pattern in this project. We have done scene identification with k-means clustering as well as without k-means clustering. We found that in both the cases as the number

of patches increases the accuracy also increases but it becomes constant after certain number of patches.

# References

- [1] T. Deselaers, D. Keysers, and H. Ney. *Discriminative training for object recognition using image patches*. In Proc. of IEEE CVPR 2005, pages 157–162, 2005.
- [2] Joo-Hwee Lim, Jean-Pierre Chevallet, Sheng Gao. *Scene Identification Using Discriminative Patterns*. 21 Heng Mui Keng Terrace, Singapore 119613.
- [3] F. Li and P. Perona. *A bayesian hierarchical model for learning natural scene categories*. In Proc. of IEEE CVPR 2005, 2005.
- [4] Dong ping Tian, Baoji, Shaanx. *A Review on Image Feature Extraction and Representation Techniques*. International Journal of Multimedia and Ubiquitous Engineering Vol. 8, No. 4, July, 2013.
- [5] Dengsheng Zhang n, Md.MonirulIslam, Guojun Lu. *A review on automatic image annotation techniques*. Pattern Recognition 45(2012)346–362.
- [6] Plinio Moreno, Alexandre Bernardino, and Jose Santos-Victor. *Gabor Parameter Selection for Local Feature Detection*. 2nd IBPRIA, Estoril, Portugal, June 7-9, 2005.